Regression Analysis: Model Application and Multicollinearity

TOTAL POINTS 10

1. For questions 1-5, download the file 'Sales by Territory.xlsx', which provides data in the following categories: Sales Revenue (in $), Territory, Quantity of Orders, and Number of Sales Calls, from a sample of size 38. 1 point

Sales by Territory.xlsx

Please run a multiple regression with Sales Revenue as your dependent variable to estimate the coefficients for each of the independent variables; remember, for categorical variables, you will need to create dummy or indicator variables. Please use "West" Territory as your reference variable, and assume an alpha of .05.

Which of the following variables are statistically significant according to the p-values?

- [x] Number of Sales Calls
- [x] Quantity of Orders
- [ ] Intercept

2. What is the value of your Y variable when all X variables are zero? That is, what is the value of 'Sales Revenue' when all X variables are zero. Round your answer to a single decimal. 1 point

-2815.5

3. For Territory "South", what is the coefficient, rounded to two decimal places? 1 point

-12240.75

4. Based on the p-value of Territory "South", which of the following statements are true? 1 point

- [ ] The variable is solely there to fit the model.
- [●] Sales in "South" territory are less than sales in "West" territory, all other variables held at the same level.
- [●] The coefficient for Territory "South" is statistically different from Territory "West", all other variables remaining at the same level.
- [ ] There would be no managerial significance.

5. Please estimate what Sales Revenue would be for a salesperson covering the "South" Territory, making 28 Sales Calls, and taking 200 Orders, rounding the answer to two decimal points. 1 point

43797.17

6. Which of the following could be a sign that a regression model has multicollinearity issues? Please mark all that apply 1 point

- [ ] The independent variables have low p-values, yet the overall fit of the model is high.
- [x] The independent variables have high p-values, yet the overall fit of the model is high.
- [ ] The independent variables have low p-values, yet the overall fit of the model is low.
- [x] The signs on some of the coefficients are contrary to common sense.

7. Which of the following statements are false regarding multicollinearity in a regression model? 1 point

- [x] Multicollinearity is always a problem; to correct, one should remove X variables causing high correlation.
- [ ] Multicollinearity matters more when interpreting coefficient impacts; to correct, one should remove X variables causing high correlation.
- [x] Multicollinearity should always be corrected when one is using a model for predictive purposes.
- [ ] Multicollinearity may not need to be corrected for if one is only utilizing a model for predictive purposes.

8. What is the R-Square for this model, rounding to two decimal points? 1 point

0.63

9. Within our regression, we notice that some of our independent variables have statistically insignificant p-values, yet as calculated above, the R-Square for the model is okay. This may imply some multicollinearity. Please calculate the correlation between "Number of Sales Calls" and "Quantity of Orders", rounding to two decimal points. 1 point

0.76

10. Does your model suffer from multicollinearity? Please mark all that apply 1 point

- [x] Most likely, multicollinearity is not an issue in the regression model.

- [ ] Low correlation drives multicollinearity in this model.
- [ ] High correlation drives multicollinearity in this model.
- [ ] Most likely, multicollinearity is an issue in the regression model.

- [x] I, **Nitish Malpotra**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.

Learn more about Coursera's Honor Code

Save Submit